Maximum Likelihood Estimation for Multiscale Ornstein-Uhlenbeck Processes

Anastasia Papavasiliou *
Department of Statistics
University of Warwick, UK
and
Fan Zhang †
School of Banking and Finance
University of International Business and Economics, China.

September 11, 2017

Abstract

We study the problem of estimating the parameters of an Ornstein-Uhlenbeck (OU) process that is the coarse-grained limit of a multiscale system of OU processes, given data from the multiscale system. We consider both the averaging and homogenization cases and both drift and diffusion coefficients. By restricting ourselves to the OU system, we are able to substantially improve the results in [23, 21] and provide some intuition of what to expect in the general case. In particular, in the homogenisation case we derive optimal rates of sub-sampling, proving the conjecture in [23].

Keywords : multiscale diffusions, Ornstein-Uhlenbeck process, parameter estimation, maximum likelihood, subsampling.

1 Introduction

A necessary step in statistical modelling is to fit the chosen model to the data by inferring the value of the unknown parameters. In the case of stochastic differential equations (SDE), this is a well studied problem [6, 16, 24]. However, quite often, data actually comes from a multiscale SDE whilst we want to model its coarse-grain approximation. This phenomenon has been observed in many applications, ranging from econometrics [1, 2, 20] to chemical engineering [5] and molecular dynamics [23]. In this paper, we study how this inconsistency between the coarse-grained model that we fit and the microscopic dynamics from which the data is generated affects the estimation problem.

The problem of estimating the drift and variance parameters of an Ornstein-Uhlenbeck (OU) process is a standard one. Statistical inference for diffusions is a well-developed

*Email: A.Papavasiliou@warwick.ac.uk
†Email: f.zhang.2@warwick.ac.uk

area and the remaining challenges mainly concern the exact computation of the likelihood. In the case of the OU process, computing the exact likelihood is straight forward.

In the context of the OU process, the problem has been extended to the case where the differential equation is not exactly a diffusion. For example, in [15], the authors studied properties of usual drift estimator for scalar OU processes driven by fractional Brownian motion. Further more detailed studies extended the results to stronger consistency and asymptotics under various assumption: such as in [9], authors discussed the asymptotics of the drift estimators for an OU process driven by fractional Gaussian processes, sub-fractional and bi-fractional Brownian motions, with infinite observation with Hurst parameter $H \in (0, 1)$; and [18] studied strong consistency and asymptotic normality for the usual drift estimator for infinite dimensional fractional OU process, under the assumption that Hurst parameter $H \geq 1/2$.

Another extension the problem was into considering the observation process as component of a multiscale process, converging in some limit to an OU process. This is the problem we are interested in. It has also been discussed in several papers. In [3, 4], the authors compute the bias between the estimators corresponding to multiscale and approximate Ornstein-Uhlenbeck (OU) process, as a function of the subsampling step size δ and the scale factor ϵ. However, their approach is somewhat ad-hoc and limited to scalar systems.

We consider the case where the multiscale system is an OU process, where the averaging and homogenization principles still hold. We look at the MLE estimators of both the drift and diffusion coefficients of the limiting system and study their properties in the case when data comes from the multiscale system. In the averaging case (section 2), we show that the estimators are consistent and asymptotically normal. However, the homogenisation case (section 3) is much more complicated as estimators are not consistent. To construct consistent estimators, one needs to subsample the data. We show that the estimators will be consistent in that case. However, proving asymptotic normality is much more involved and beyond the scopes of this paper. Our approach is similar to that in [23, 21]. In the first, the authors study a similar problem for slightly more general models but get weaker results. In the latter, the authors study the behaviour of the drift likelihood of the limiting system for data coming from the multiscale system.

Finally, let us note that MLE estimators for diffusions are known to have practical limitations. In particular, the drift estimator requires a long horizon to achieve satisfactory precision while the diffusion estimator needs very fine scale data. In the homogenisation case that we study, there is the additional issue of subsampling the data at a sampling rate that depends on normally unknown separation of scales variable ϵ. However, we believe that the behaviour that we observe is inherent to the problem of mismatch between the model and data and not the estimators and thus have chosen a set-up where analysis can be done in detail.

We present the exact set-up and main ideas below. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t>0}, \mathbb{P})$ be a filtered probability space, and U, V be two independent Brownian motions defined on this space. We consider multiscale systems of SDEs of the form

$$\frac{dx_t^\epsilon}{dt} = a_{11}x_t^\epsilon + a_{12}y_t^\epsilon + \sqrt{q_1}\frac{dU_t}{dt}, \quad x^\epsilon = x_0 \tag{1a}$$

$$\frac{dy_t^\epsilon}{dt} = \frac{1}{\epsilon}\left(a_{21}x_t^\epsilon + a_{22}y_t^\epsilon\right) + \sqrt{\frac{q_2}{\epsilon}}\frac{dV_t}{dt}, \quad y^\epsilon = y_0 \tag{1b}$$

or

$$\frac{dx_t^\epsilon}{dt} = \frac{1}{\epsilon}\left(a_{11}x_t^\epsilon + a_{12}y_t^\epsilon\right) + \left(a_{13}x_t^\epsilon + a_{14}y_t^\epsilon\right) + \sqrt{q_1}\frac{dU_t}{dt}, \;\; x^\epsilon = x_0 \tag{2a}$$

$$\frac{dy_t^\epsilon}{dt} = \frac{1}{\epsilon^2}\left(a_{21}x_t^\epsilon + a_{22}y_t^\epsilon\right) + \sqrt{\frac{q_2}{\epsilon^2}}\frac{dV_t}{dt}, \;\; y^\epsilon = y_0 \tag{2b}$$

where x_0, y_0 are random variables, $(a_{11}, a_{12}, a_{13}, a_{14}, a_{21}, a_{22})$ are real constants that will be required to satisfied certain relationships to be specified later which guarantee ergodicity and q_1, q_2 are positive real constants. The variable $\epsilon > 0$ denotes scale separation and we will consider the behaviour of the above system in the limit $\epsilon \to 0$. We refer to equations (1) and (2) as the averaging and homogenization case, respectively and we denote by $\rho_\epsilon(x, y)$ their invariant distribution, when this exists. In both cases and under certain conditions, $(x_t^\epsilon)_{0\leq t\leq T}$ converges as $\epsilon \to 0$ to the solution of

$$\frac{dX_t}{dt} = \tilde{a}X_t + \sqrt{\sigma}\frac{dW_t}{dt}, \tag{3}$$

for appropriate $\tilde{a} \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ and in a way to be made precise later, for W also Brownian motion defined on the probability space. Our goal will be to estimate a and σ, assuming that we continuously observe $(x_t^\epsilon)_{0\leq t\leq T}$ from (1) or (2). It is a well known result (see [6, 19]) that, given $(X_t)_{0\leq t\leq T}$, the maximum likelihood estimators for a is

$$\hat{a}_T = \left(\int_0^T X_t dX_t\right)\left(\int_0^T X_t^2 dt\right)^{-1}. \tag{4}$$

If $(X_t)_{0\leq t\leq T}$ is discretely observed, then the estimator of σ is the discretised Quadratic Variation

$$\hat{\sigma}_\delta = \frac{1}{T}\sum_{n=0}^{N-1}\left(X_{(n+1)\delta} - X_{n\delta}\right)^2 \tag{5}$$

which converges in L_2 to σ as $\delta \to 0$. Our approach will be to still use the estimators defined in (4) and (5), replacing X_t by its x_t^ϵ approximation coming from the multiscale model and then studying their asymptotic properties. In section 2, we discuss the averaging case while in section 3 we study the homogenization case.

We shall discuss problems in scalars for simplicity of notation and writing. However, the conclusions can easily be extended to finite dimensions. We will use c to denote an arbitrary constant which can vary from occurrence to occurrence. Also, for the sake of simplicity we will sometimes write x_n^ϵ (or y_n^ϵ, X_n) instead of $x_{n\delta}^\epsilon$ (resp. $y_{n\delta}^\epsilon$, $X_{n\delta}$). Finally, note that the transpose of an arbitrary matrix A is denoted by A^*.

2 Averaging

We consider the system of stochastic differential equations described by (1) (averaging case), where $(x_t^\epsilon, y_t^\epsilon) \in \mathcal{X} \times \mathcal{Y}$. We may take $\mathcal{X}$ and $\mathcal{Y}$ as either $\mathbb{R}$ or $\mathbb{T}$. We make the following assumptions:

Assumptions 2.1.
Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t>0}, \mathbb{P})$ be a filtered probability space. We assume that

(i) $(U_t, V_t)_{t\geq 0}$ are independent Brownian motions;

(*ii*) q_1, q_2 *are positive real constants;*

(*iii*) $0 < \epsilon \ll 1$ *is the scale separation variable;*

(*iv*) $a_{22} < 0$ *and* $a_{11} < a_{12}a_{22}^{-1}a_{21}$*;*

(*v*) x_0 *and* y_0 *are random variables, independent of* U, V *and* $\mathbb{E}\left(x_0^2 + y_0^2\right) < \infty$.

These assumptions guarantee the ergodicity of system (1). In this case, the averaging limit of the system is given by the following equation (see [14]):

$$\frac{dX_t}{dt} = \tilde{a}X_t + \sqrt{q_1}\frac{dU_t}{dt} \tag{6}$$

where:

$$\tilde{a} = a_{11} - a_{12}a_{22}^{-1}a_{21} \tag{7}$$

2.1 The Paths

In this section, we show that $(x_t^\epsilon)_{0\le t\le T}$ defined in (1) converges in a strong sense to the solution $X_{0\le t\le T}$ of (6). Our result extends that of [22] (Theorem 17.1) where the state space $\mathcal{X}$ is restricted to $\mathbb{T}$ and the averaging equation is deterministic. Assuming that the system is an OU process, the domain can be extended to $\mathbb{R}$ and the averaging equation can be stochastic. We prove the following lemma first:

Lemma 2.2. *Suppose that* $(x_t^\epsilon, y_t^\epsilon)_{0\le t\le T}$ *solves* (1) *and Assumptions 2.1 are satisfied. Then, for finite* $T > 0$ *and* ϵ *small,*

$$\mathbb{E}\sup_{0\le t\le T}\left[(x_t^\epsilon)^2 + (y_t^\epsilon)^2\right] \approx \mathcal{O}\left(\log\left(1 + \frac{T}{\epsilon}\right)\right). \tag{8}$$

Proof. Since $(U_t)_{t\ge 0}$ and $(V_t)_{t\ge 0}$ are independent, we can rewrite (1) in vector form as

$$d\mathbf{x}_t^\epsilon = \mathbf{a}_\epsilon \mathbf{x}_t^\epsilon dt + \sqrt{\mathbf{q}_\epsilon}d\mathbf{W}_t \tag{9}$$

where

$$\mathbf{x}_t^\epsilon = \begin{pmatrix} x_t^\epsilon \\ y_t^\epsilon \end{pmatrix}, \mathbf{a}_\epsilon = \begin{pmatrix} a_{11} & a_{12} \\ \frac{1}{\epsilon}a_{21} & \frac{1}{\epsilon}a_{22} \end{pmatrix}, \mathbf{q}_\epsilon = \begin{pmatrix} q_1 & 0 \\ 0 & \frac{q_2}{\epsilon} \end{pmatrix}$$

and $\mathbf{W_t} = (U_t, V_t)$ is two-dimensional Brownian motion. Given the form of $\mathbf{a}_\epsilon$, it is an easy exercise to show that its eigenvalues will be of order $\mathcal{O}(1)$ and $\mathcal{O}(\frac{1}{\epsilon})$. Therefore, we define the eigenvalue decomposition of $\mathbf{a}_\epsilon$ as

$$\mathbf{a}_\epsilon = P_\epsilon D_\epsilon P_\epsilon^{-1} \text{ with } D_\epsilon = \begin{pmatrix} \lambda_1(\epsilon) & 0 \\ 0 & \frac{1}{\epsilon}\lambda_2(\epsilon) \end{pmatrix},$$

where $\lambda_1(\epsilon), \lambda_2(\epsilon)$ are both of order $\mathcal{O}(1)$. Again, it is not hard to see that if $(p_1^\epsilon, p_2^\epsilon)$ is an eigenvector, $\mathcal{O}(p_1^\epsilon) = \mathcal{O}(\lambda(\epsilon)_i^{-1}p_2^\epsilon)$, for $i = 1, 2$ depending on the corresponding eigenvector. So, for the eigenvector corresponding to eigenvalue of order $\mathcal{O}(1)$, all elements of the eigenvector will also be of order $\mathcal{O}(1)$ while for the eigenvector corresponding to eigenvalue of order $\mathcal{O}(1/\epsilon)$, we will have that $p_1^\epsilon \sim \mathcal{O}(1)$ and $p_2^\epsilon \sim \mathcal{O}(\epsilon)$.

Now, let us define $\Sigma_\epsilon = P_\epsilon^{-1}\mathbf{q}_\epsilon(P_\epsilon^{-1})^*$. It follows that

$$\Sigma_\epsilon = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(1) & \mathcal{O}(1/\epsilon) \end{pmatrix}$$

We apply a linear transformation to the system of equations (9) so that the drift matrix becomes diagonal. It follows from [10] that

$$\mathbb{E}\left(\sup_{0\le t\le T}\|\mathbf{x}_t^\epsilon\|^2\right)\le C\frac{\log\left(1+\max_i(|(D_\epsilon)_{ii}|)T\right)}{\min_i(|(D_\epsilon)_{ii}/(\Sigma_\epsilon)_{ii}|)}, i\in\{1,2\}.$$

Since the diagonal elements of D_ϵ and Σ_ϵ are of the same order and $\max_i|(D_\epsilon)_{ii}| = \mathcal{O}(\frac{1}{\epsilon})$, we have

$$\mathbb{E}\left(\sup_{0\le t\le T}\|\mathbf{x}_t^\epsilon\|^2\right)=\mathcal{O}\left(\log(1+T/\epsilon)\right).$$

The result follows by expanding the vector norm. □

Theorem 2.3. *Let Assumptions 2.1 hold for system* (1)*. Suppose that* $(x_t^\epsilon, y_t^\epsilon)_{0\le t\le T}$ *and* $(X_t)_{0\le t\le T}$ *are two solutions of* (1) *and* (6) *respectively, corresponding to the same realization of the* $(U_t)_{t\ge0}$ *process and* $x_0 = X_0$*. Then,* $(x_t^\epsilon)_{0\le t\le T}$ *converges to* $(X_t)_{0\le t\le T}$ *in* $L^2\left(\Omega, C([0,T],\mathcal{X})\right)$*. More specifically,*

$$\mathbb{E}\sup_{0\le t\le T}(x_t^\epsilon - X_t)^2\le c\left(\epsilon^2 log\left(\frac{T}{\epsilon}\right)+\epsilon T\right)e^{T^2}.$$

Note that when the time horizon T *is fixed finite, the above bound can be simplified to*

$$\mathbb{E}\sup_{0\le t\le T}(x_t^\epsilon - X_t)^2=\mathcal{O}(\epsilon).$$

Proof. The first step in the proof will be to expand the slow variable x_t^ϵ in (1a) in terms of ϵ. In the OU case, we can get the expansion directly by solving for y_t^ϵ in (1b) and using the answer to replace y_t^ϵ in (1a). Note that a more general approach that can be applied to nonlinear systems is to use Poisson equations (see [22]).

Solving (1b) for y_t^ϵ gives

$$y_t^\epsilon = -a_{22}^{-1}a_{21}x_t^\epsilon - \sqrt{\epsilon}a_{22}^{-1}\sqrt{q_2}\frac{dV_t}{dt}+\epsilon a_{22}^{-1}\frac{dy_t^\epsilon}{dt} \tag{10}$$

and replacing y_t^ϵ by (10) in (1a) gives

$$dx_t^\epsilon = \tilde{a}x_t^\epsilon dt+\sqrt{q_1}dU_t+\sqrt{\epsilon}a_{12}a_{22}^{-1}\sqrt{q_2}dV_t+\epsilon a_{12}a_{22}^{-1}dy_t^\epsilon, \tag{11}$$

where $\tilde{a}$ is defined in (7). It follows that

$$x_t^\epsilon = x_0+\int_0^t\tilde{a}x_s^\epsilon ds+\sqrt{q_1}U_t+\sqrt{\epsilon}a_{12}a_{22}^{-1}\sqrt{q_2}V_t+\epsilon a_{12}a_{22}^{-1}(y_t^\epsilon - y_0).$$

Also, from the averaged equation (6), we get

$$X_t = X_0+\int_0^t\tilde{a}X_s ds+\sqrt{q_1}U_t.$$

Let $e(\epsilon)_t = x_t^\epsilon - X_t$. By assumption, $e(\epsilon)_0 = 0$ and

$$e(\epsilon)_t = \int_0^t\tilde{a}e(\epsilon)_s ds+\sqrt{\epsilon}a_{12}a_{22}^{-1}\sqrt{q_2}V_t+\epsilon a_{12}a_{22}^{-1}(y_t^\epsilon - y_0). \tag{12}$$

Then,

$$e(\epsilon)_t^2 \le 3\left(\left(\tilde{a}\int_0^t e(\epsilon)_s ds\right)^2 + \epsilon\left(a_{12}a_{22}^{-1}\right)^2 V_t^2 + \epsilon^2\left(a_{12}a_{22}^{-1}\right)^2 (y_t^\epsilon - y_0)^2\right) .$$

Apply Lemma 2.2, the Burkholder-Davis-Gundy inequality [22], Hölder inequality and Itô isometry on (12), we get

$$\begin{aligned}\mathbb{E}\left(\sup_{0\le t\le T} e(\epsilon)_t^2\right) &\le c\left(T\int_0^T \mathbb{E}e(\epsilon)_s^2 ds + \epsilon^2\log(\frac{T}{\epsilon}) + \epsilon T\right)\\ &\le c\left(\epsilon^2\log(\frac{T}{\epsilon}) + \epsilon T + T\int_0^T \mathbb{E}\sup_{0\le u\le s} e(\epsilon)_u^2 ds\right).\end{aligned}$$

By Gronwall's inequality [22], we deduce that

$$\mathbb{E}\left(\sup_{0\le t\le T}(e(\epsilon)_t)^2\right) \le c(\epsilon^2\log(\frac{T}{\epsilon}) + \epsilon T)e^{T^2}.$$

□

2.2 The Drift Estimator

Suppose that we want to estimate the drift of the process $(X_t)_{0\le t\le T}$ described by (6) but we only observe a solution $(x_t^\epsilon)_{0\le t\le T}$ of (1a) for some $\epsilon > 0$. According to the previous theorem, $(x_t^\epsilon)_{0\le t\le T}$ is a good approximation of $(X_t)_{0\le t\le T}$, so we replace $(X_t)_{0\le t\le T}$ in the formula of the MLE (4) by $(x_t^\epsilon)_{0\le t\le T}$. In the following theorem, we show that the error we will be making is insignificant, in a sense to be made precise.

Theorem 2.4. *Suppose that $(x_t^\epsilon, y_t^\epsilon)_{0\le t\le T}$ solves system* (1)*, satisfying Assumptions 2.1. Let $\hat{a}_T^\epsilon$ be the estimate we get by replacing X_t in* (4) *by x_t^ϵ, i.e.*

$$\hat{a}_T^\epsilon = \left(\int_0^T x_t^\epsilon dx_t^\epsilon\right)\left(\int_0^T (x_t^\epsilon)^2 dt\right)^{-1} . \tag{13}$$

Then,

$$\lim_{\epsilon\to 0}\lim_{T\to\infty}\mathbb{E}(\hat{a}_T^\epsilon - \tilde{a})^2 = 0 ,$$

for $\tilde{a}$ given by (7).

Proof. We define

$$I_1(T) = \frac{1}{T}\int_0^T x_t^\epsilon dx_t^\epsilon \text{ and } I_2(T) = \frac{1}{T}\int_0^T (x_t^\epsilon)^2 dt.$$

By ergodicity, which is guaranteed by Assumptions 2.1 *(iii) and (iv)*

$$\lim_{T\to\infty} I_2(T) = \mathbb{E}\left((x_\infty^\epsilon)^2\right) = C \ne 0 \; a.s.,$$

where x_∞^ϵ is a random variable distributed according to the marginal of invariant distribution ρ_ϵ of system (1). This is an $\mathcal{O}(1)$ non-zero constant, for all values of the parameters under assumptions 2.1. Using the (11) expansion of dx_t^ϵ in terms of ϵ, we get

$$
\begin{aligned}
I_1(T) &= \tilde{a} I_2(T) + \qquad (14)\\
&+ \sqrt{q_1}\frac{1}{T}\int_0^T x_t^\epsilon dU_t + \sqrt{\epsilon} a_{12} a_{22}^{-1}\sqrt{q_2}\frac{1}{T}\int_0^T x_t^\epsilon dV_t + a_{12}a_{22}^{-1}\frac{1}{T}\int_0^T x_t^\epsilon \epsilon dy_t^\epsilon.
\end{aligned}
$$

From Itô isometry and ergodicity, we directly get that

$$\mathbb{E}\left(\sqrt{q_1}\frac{1}{T}\int_0^T x_t^\epsilon dU_t\right)^2 = q_1\frac{1}{T^2}\int_0^T \mathbb{E}(x_t^\epsilon)^2 dt = \frac{c}{T}$$

and similarly,

$$\mathbb{E}\left(\sqrt{\epsilon} a_{12} a_{22}^{-1}\sqrt{q_2}\frac{1}{T}\int_0^T x_t^\epsilon dV_t\right)^2 = \frac{\epsilon c}{T}.$$

Finally, using (1b), we break the last term of (14) further into

$$\frac{1}{T}\int_0^T x_t^\epsilon \epsilon dy_t^\epsilon = -\frac{1}{T}\int_0^T x_t^\epsilon (a_{21}x_t^\epsilon + a_{22}y_t^\epsilon) dt - \frac{\sqrt{\epsilon q_2}}{T}\int_0^T x_t^\epsilon dV_t$$

As before, we see that the last term will be of order $\mathcal{O}\left(\frac{\epsilon}{T}\right)$. By ergodicity, the first term converges in L^2 as $T \to \infty$

$$-\frac{1}{T}\int_0^T x_t^\epsilon (a_{21}x_t^\epsilon + a_{22}y_t^\epsilon) dt \to \mathbb{E}\left(x_\infty^\epsilon (a_{21}x_\infty^\epsilon + a_{22}y_\infty^\epsilon)\right),$$

where, as before, $(x_\infty^\epsilon, y_\infty^\epsilon)$ are random variable distributed according to the invariant distribution ρ_ϵ of system (1). We write the above expectation as

$$\mathbb{E}\left(x_\infty^\epsilon (a_{21}x_\infty^\epsilon + a_{22}y_\infty^\epsilon)\right) = \mathbb{E}\left(x_\infty^\epsilon \mathbb{E}\left((a_{21}x_\infty^\epsilon + a_{22}y_\infty^\epsilon)|x_\infty^\epsilon\right)\right).$$

Clearly, the limit of ρ_ϵ conditioned on x_∞^ϵ is a normal distribution with mean $-a_{22}^{-1}a_{21}x_\infty^\epsilon$. Thus, we see that

$$\lim_{\epsilon\to 0}\mathbb{E}\left(x_\infty^\epsilon (a_{21}x_\infty^\epsilon + a_{22}y_\infty^\epsilon)\right) = 0.$$

Putting everything together, we see that

$$\lim_{\epsilon\to 0}\lim_{T\to\infty}(I_1 - \tilde{a} I_2) = 0 \quad \text{in} \quad L^2$$

Since the denominator I_2 of $\hat{a}_T^\epsilon$ converges almost surely, the result follows. ☐

2.2.1 Asymptotic Normality for the Drift Estimator

We extend the proof of Theorem 2.4 to prove asymptotic normality for the estimator $\hat{a}_T^\epsilon$. We will show that

$$\sqrt{T}\left(\hat{a}_T^\epsilon - \tilde{a} + a_{12}\mathbb{E}\left(x_\infty^\epsilon(a_{22}^{-1}a_{21}x_\infty^\epsilon + y_\infty^\epsilon)\right)\right) \to \mathcal{N}\left(0, \sigma_\epsilon^2\right)$$

in distribution, as $T \to \infty$ and compute the limit of σ_ϵ^2 as $\epsilon \to 0$. We start with expansion (14). First we apply the Central Limit Theorem to the martingales (see [11]). We find that

$$\sqrt{T}\left(\sqrt{q_1}\frac{1}{T}\int_0^T x_t^\epsilon dU_t\right) \to \mathcal{N}\left(0, \sigma(1,1)_\epsilon^2\right) \quad \text{as } T \to \infty$$

where

$$\sigma(1,1)^2_\epsilon = q_1 \mathbb{E}[(x^\epsilon_\infty)^2]$$

and

$$\sqrt{T}\left(\sqrt{\epsilon} a_{12} a_{22}^{-1} \sqrt{q_2} \frac{1}{T} \int_0^T x^\epsilon_t dV_t\right) \to \mathcal{N}\left(0, \sigma(1,2)^2_\epsilon\right) \quad \text{as } T \to \infty$$

where

$$\sigma(1,2)^2_\epsilon = \epsilon q_2 (a_{12} a_{22}^{-1})^2 \mathbb{E}\left((x^\epsilon_\infty)^2\right).$$

As before, we further expand the last component of the expansion (14) to

$$J_1 = -\frac{a_{12} a_{22}^{-1}}{T} \int_0^T (a_{21} x_t^2 + a_{22} x_t y_t) dt \;\text{ and }\; J_2 = -\frac{a_{12} a_{22}^{-1} \sqrt{\epsilon q_2}}{T} \int_0^T x_t dV_t.$$

Once again, we apply the Central Limit Theorem for martingales to J_2 and we find

$$\sqrt{T} J_2 \to \mathcal{N}\left(0, \sigma(2,2)^2_\epsilon\right) \quad \text{as } T \to \infty$$

where

$$\sigma(2,2)^2_\epsilon = \epsilon (a_{21} a_{22}^{-1})^2 q_2 \mathbb{E}\left((x^\epsilon_\infty)^2\right).$$

Finally, we apply the Central Limit Theorem for functionals of ergodic Markov Chains to J_1 (see [7]). We get

$$\sqrt{T}\left(J_1 + a_{12} \mathbb{E}\left(x^\epsilon_\infty (a_{22}^{-1} a_{21} x^\epsilon_\infty + y^\epsilon_\infty)\right)\right) \to \mathcal{N}\left(0, \sigma(2,1)^2_\epsilon\right)$$

as $T \to \infty$, where with

$$\sigma(2,1)^2_\epsilon = \mathrm{Var}\left(\xi(x^\epsilon_\infty, y^\epsilon_\infty)\right) + 2 \int_0^\infty \mathrm{Cov}\left(\xi(x_0, y_0), \xi(x^\epsilon_t, y^\epsilon_t)\right) dt,$$

where

$$\xi(x,y) = -\left(a_{12} a_{22}^{-1} a_{21} x^2 + a_{12} x y\right).$$

Putting everything together, we get that as $T \to \infty$,

$$\sqrt{T}\left(I_1(T) - \tilde{a} I_2(T)\right) \to Z(1,1)_\epsilon + Z(1,2)_\epsilon + Z(2,1)_\epsilon + Z(2,2)_\epsilon,$$

in law, where $Z(i,j) \sim \mathcal{N}(0, \sigma(i,j)^2_\epsilon)$, for $i,j = 1,2$. Finally, we note that the denominator I_2 converges almost surely as $T \to \infty$ to $\mathbb{E}\left((x^\epsilon_\infty)^2\right)$. It follows from Slutsky's theorem that as $T \to \infty$,

$$\sqrt{T}\left(\hat{a}^\epsilon_T - \tilde{a} + a_{12} \mathbb{E}\left(x^\epsilon_\infty (a_{22}^{-1} a_{21} x^\epsilon_\infty + y^\epsilon_\infty)\right)\right) \to \mathcal{N}(0, \sigma^2_\epsilon),$$

where

$$\sigma^2_\epsilon = \frac{\mathbb{E}(Z(1,1)_\epsilon + Z(1,2)_\epsilon + Z(2,1)_\epsilon + Z(2,2)_\epsilon)^2}{\mathbb{E}\left((x^\epsilon_\infty)^2\right)^2}.$$

It remains to compute $\lim_{\epsilon \to 0} \sigma^2_\epsilon$. We have already seen that $\sigma(1,2)^2_\epsilon \sim \mathcal{O}(\epsilon)$ and $\sigma(2,2)^2_\epsilon \sim \mathcal{O}(\epsilon)$, so we don't expect $Z(1,2)_\epsilon$ and $Z(2,2)_\epsilon$ to contribute to the limit. Also,

$$\sigma(1,1)^2_\epsilon = q_1 \mathbb{E}\left((x^\epsilon_\infty)^2\right) \to q_1 \mathbb{E}\left(X^2_\infty\right) = -\frac{q_1^2}{2\tilde{a}},$$

where X_∞ is distributed according to the invariant distribution of system (6). To compute $\lim_{\epsilon \to 0} \sigma(2,1)^2_\epsilon$, we set $\tilde{y}(x,y) = a_{22}^{-1} a_{21} x + y$. Then, $(x^\epsilon_t, \tilde{y}(x^\epsilon_t, y^\epsilon_t))$ is also an ergodic process with invariant distribution $\tilde{\rho}_\epsilon$ that converges as $\epsilon \to 0$ to $\mathcal{N}(0, \frac{q_1}{2\tilde{a}}) \otimes \mathcal{N}(0, \frac{q_2}{2a_{22}})$. Since $\xi(x,y) = -a_{21}x \cdot \tilde{y}(x,y)$, it follows that

$$\lim_{\epsilon\to 0} \mathrm{Var}\,(\xi(x^\epsilon_\infty, y^\epsilon_\infty)) = a_{12}^2 \frac{q_1}{2\tilde{a}} \frac{q_2}{2a_{22}}.$$

In addition, as $\epsilon \to 0$, the process $\tilde{y}(x^\epsilon_t, y^\epsilon_t)$ decorrelates exponentially fast. Thus

$$\lim_{\epsilon\to 0} \mathrm{Cov}\,(\xi(x_0,y_0), \xi(x^\epsilon_t, y^\epsilon_t)) = a_{12}^2 \mathrm{Cov}\,(X_0, X_t) \lim_{\epsilon\to 0} \mathrm{Cov}\,(\tilde{y}(x_0,y_0), \tilde{y}(x^\epsilon_t, y^\epsilon_t)) \equiv 0$$

for all $t \geq 0$. As $t \to \infty$, the process $(x^\epsilon_t, \tilde{y}(x^\epsilon_t, y^\epsilon_t))$ also converges exponentially fast to a mean-zero Gaussian distribution and thus the integral with respect to t is finite. We conclude that the second term of $\sigma(2,1)^2_\epsilon$ disappears as $\epsilon \to 0$ and thus

$$\lim_{\epsilon\to 0} \mathbb{E}(Z(2,1)^2_\epsilon) = a_{12}^2 \frac{q_1 q_2}{4\tilde{a}a_{22}}.$$

Finally, we show that

$$\lim_{\epsilon\to 0} \mathbb{E}(Z(2,1)_\epsilon Z(1,1)_\epsilon) = 0.$$

Clearly, $Z(1,1)_\epsilon$ is independent of $\tilde{y}(x^\epsilon_t, y^\epsilon_t)$ in the limit, since it only depends on the processes $\{x^\epsilon_t\}_{t>0}$ and $\{U_t\}_{t>0}$. So,

$$\lim_{\epsilon\to 0} \mathbb{E}(Z(2,1)_\epsilon Z(1,1)_\epsilon) = \lim_{\epsilon\to 0} \mathbb{E}\left(\mathbb{E}(Z(2,1)_\epsilon Z(1,1)_\epsilon | x^\epsilon_t,\ t>0)\right)$$

and

$$\lim_{\epsilon\to 0} \mathbb{E}\left(\mathbb{E}(Z(2,1) | x^\epsilon_t,\ t>0)\right) = 0$$

for the same reasons as above. Thus

$$\lim_{\epsilon\to 0} \sigma^2_\epsilon = \frac{4\tilde{a}^2}{q_1^2}\left(-\frac{q_1^2}{2\tilde{a}} + a_{12}^2 \frac{q_1 q_2}{4\tilde{a}a_{22}}\right).$$

We have proved the following

Theorem 2.5. *Suppose that $(x^\epsilon_t, y^\epsilon_t)_{0\leq t\leq T}$ is a solution of system* (1) *satisfying Assumptions 2.1. Let $\hat{a}^\epsilon_T$ be as in* (13)*. Then,*

$$\sqrt{T}\,(\hat{a}^\epsilon_T - \tilde{a} - \mu_\epsilon) \to \mathcal{N}(0, \sigma^2_\epsilon),$$

where

$$\mu_\epsilon \to 0 \quad \text{and} \quad \sigma^2_\epsilon \to -2\tilde{a} + a_{12}^2 \frac{\tilde{a}q_2}{a_{22}q_1} \quad \text{as} \quad \epsilon \to 0.$$

Remark 2.6. *Note that in the case where the data comes from the multiscale limit and for $\epsilon \to 0$, the asymptotic variance of the drift MLE is larger than that the asymptotic variance of the drift estimator where there is no misfit between model and data. The asymptotic variance of the drift MLE with data coming from the averaged system,* (4)*, is given by $-2\tilde{a}$.*

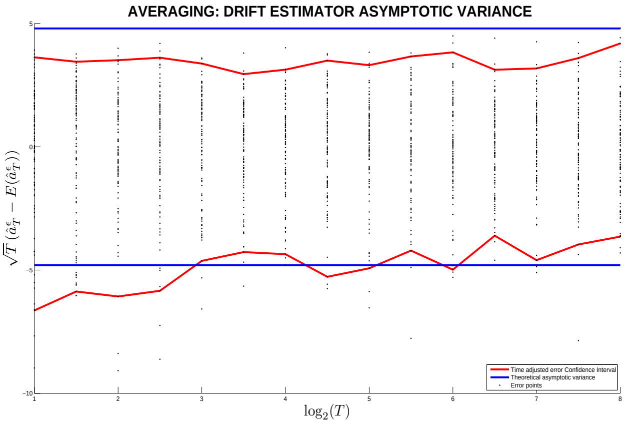

Figure 1: The asymptotic variance of the drift estimator is constructed by plotting the distribution of the time adjusted errors $\sqrt{T}\,(\hat{a}^\epsilon_T - \mathbb{E}(\hat{a}^\epsilon_T))$ for the following choice of parameters: $a_{11} = a_{21} = a_{22} = -1$, $a_{12} = 1$, $q_1 = q_2 = 2$ and $\epsilon = 2^{-9}$. Also, T is sampled from 2^1 to 2^8. The blue lines are the theoretical bounds as described in Theorem 2.5, the red lines are the 2.5 and 97.5 percentiles from the simulated samples.

2.3 The Diffusion Estimator

Suppose that we want to estimate the diffusion parameter of the process $(X_t)_{0\le t\le T}$ described by (6) but we only observe a solution $(x^\epsilon_t)_{0\le t\le T}$ of (1a). As before, we replace X_t in the formula of the MLE (5) by x^ϵ_t. The following theorem states that the estimator is still consistent in the limit.

Theorem 2.7. *Suppose that $(x^\epsilon_t, y^\epsilon_t)_{0\le t\le T}$ is the solution of system* (1) *satisfying Assumptions 2.1. We set*

$$\hat{q}^\epsilon_\delta = \frac{1}{T}\sum_{n=0}^{N-1}\left(x^\epsilon_{(n+1)\delta} - x^\epsilon_{n\delta}\right)^2 \tag{15}$$

where $\delta \le \epsilon$ is the discretization step and $T = N\delta$ is fixed. Then, for every $\epsilon > 0$

$$\lim_{\delta\to 0}\mathbb{E}(\hat{q}^\epsilon_\delta - q_1)^2 = 0\,.$$

In addition, and in distribution

$$\delta^{-\frac{1}{2}}\,(\hat{q}^\epsilon_\delta - q_1) \xrightarrow{D} \mathcal{N}(0, \frac{2q_1^2}{T}) \text{ as } \delta \to 0\,.$$

The proof of the theorem is the standard proof of quadratic variation converging to the diffusion coefficient, as in the averaging case, the diffusion coefficient remains unaltered in the limit.

In Figure 2, we show an example of the distributions of the errors of the diffusion estimator as $\delta \to 0$.

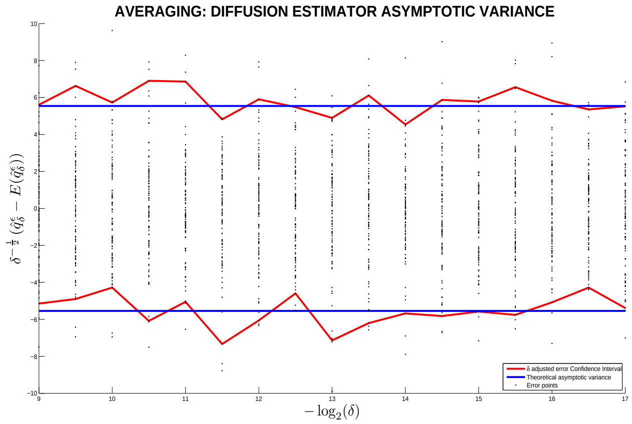

Figure 2: The asymptotic variance of the diffusion estimator is constructed by plotting the distribution of δ adjusted errors $\delta^{\frac{1}{2}}\,(\hat{q}^{\epsilon}_{\delta} - \mathbb{E}(\hat{q}^{\epsilon}_{\delta}))$ for the following choice of parameters: $a_{11} = a_{21} = a_{22} = -1$, $a_{12} = 1$, $q_1 = q_2 = 2$ and $\epsilon = 2^{-9}$. Also, δ sampled from 2^{-9} to 2^{-17}. The blue and red lines correspond to the theoretical and simulated 95% confidence intervals.

3 Homogenization

We now consider the system of stochastic differential equations described by (2), for the variables $(x^{\epsilon}_t, y^{\epsilon}_t) \in \mathcal{X} \times \mathcal{Y}$. We may take $\mathcal{X}$ and $\mathcal{Y}$ as either in $\mathbb{R}$ or $\mathbb{T}$. Our interest remains in data generated by the $(x^{\epsilon}_t)_{0\le t\le T}$ process.

Assumptions 3.1.
We assume that Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t>0}, \mathbb{P})$ be a filtered probability space. We assume that

(i) $(U_t, V_t)_{t\ge 0}$ *are independent Brownian motions;*

(ii) q_1, q_2 *are positive real constants;*

(iii) $0 < \epsilon \ll 1$ *is the scale separation variable;*

(iv) Constants $a_{11}, a_{12}, a_{13}, a_{14}, a_{21}, a_{22}$ *are all real valued and the system's drift matrix*

$$\begin{pmatrix} \frac{1}{\epsilon}a_{11} + a_{13} & \frac{1}{\epsilon}a_{12} + a_{14} \\ \frac{1}{\epsilon^2}a_{21} & \frac{1}{\epsilon^2}a_{22} \end{pmatrix}$$

only has negative real eigenvalues when ϵ is sufficiently small;

(v) $a_{21} \neq 0$*;*

(vi) x_0 *and* y_0 *are random variables, independent of* U, V*. Moreover,* $\mathbb{E}\left(x_0^2 + y_0^2\right) < \infty$*.*

Remark 3.2. *Assumption 3.1(iv) guarantees the ergodicity of the whole system* (2) *for ϵ is sufficiently small. This condition can be decomposed to a_{22} and $a_{13} - a_{14}a_{22}^{-1}a_{21}$ being negative real numbers and $a_{11} - a_{12}a_{22}^{-1}a_{21} = 0$, which ensures that the fast scale term in (2a) vanishes.*

Remark 3.3. *Assumption 3.1(v) is necessary in our setup. However, for the discussion of the case where $a_{21} = 0$ is zero, see [8].*

The corresponding homogenized equation is given by (see [14]):

$$dX_t = \tilde{a} X_t + \sqrt{\tilde{q}} dW_t \tag{16}$$

where

$$\tilde{a} = a_{13} - a_{14} a_{22}^{-1} a_{21} \tag{17}$$

and

$$\tilde{q} = q_1 + a_{12}^2 a_{22}^{-2} q_2 \tag{18}$$

and for $(W_t)_{t>0}$ Brownian motion.

Below, we will show that, similar to the averaging case, the paths of the slow process converge to the paths of the corresponding homogenized equation. However, we will see that in the limit $\epsilon \to 0$, the likelihood of the observations is different depending on whether we observe a path of the slow process generated by (2a) or the homogenized process (16) (see also [21, 22, 23]).

3.1 The Paths

The following theorem extends Theorem 18.1 in [22], which gives weak convergence of paths on $\mathbb{T}$. By limiting ourselves to the OU process, we extend the domain to $\mathbb{R}$ and prove a stronger mode of convergence.

Lemma 3.4. *Suppose that $(x_t^\epsilon, y_t^\epsilon)_{0\le t\le T}$ solves* (2) *and Assumptions 3.1 are satisfied. Then, for fixed finite $T > 0$ and small ϵ,*

$$\mathbb{E} \sup_{0\le t\le T} \left((x_t^\epsilon)^2 + (y_t^\epsilon)^2\right) = \mathcal{O}\left(\log(1 + \frac{T}{\epsilon^2})\right). \tag{19}$$

Proof. We look at the system of SDEs as,

$$d\mathbf{x}_t^\epsilon = \mathbf{a}_\epsilon \mathbf{x}_t^\epsilon dt + \sqrt{\mathbf{q}_\epsilon} dW_t \tag{20}$$

where,

$$\mathbf{x}_t^\epsilon = \begin{pmatrix} x_t^\epsilon \\ y_t^\epsilon \end{pmatrix}, \mathbf{a}_\epsilon = \begin{pmatrix} \frac{1}{\epsilon} a_{11} + a_{13} & \frac{1}{\epsilon} a_{12} + a_{14} \\ \frac{1}{\epsilon^2} a_{21} & \frac{1}{\epsilon^2} a_{22} \end{pmatrix} \text{ and } \mathbf{q}_\epsilon = \begin{pmatrix} q_1 & 0 \\ 0 & \frac{1}{\epsilon^2} q_2 \end{pmatrix}.$$

We want to characterize the magnitude of the eigenvalues of $\mathbf{a}_\epsilon$. Using existing results regarding the eigenvalues of a perturbed matrix (see [12], p. 137, Theorem 2), we find that the eigenvalues will be of order $\mathcal{O}(1)$ and $\mathcal{O}(1/\epsilon^2)$. Therefore, we can decompose $\mathbf{a}_\epsilon$ as

$$\mathbf{a}_\epsilon = P_\epsilon D_\epsilon P_\epsilon^{-1} \text{ with } D_\epsilon = \begin{pmatrix} \lambda_1(\epsilon) & 0 \\ 0 & \frac{1}{\epsilon^2} \lambda_2(\epsilon) \end{pmatrix}$$

where D_ϵ is the diagonal matrix, for which $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \in \mathbb{R}$ are diagonal entries of order $\mathcal{O}(1)$. Following exactly the same approach as in lemma 2.2, we get the result. $\square$

Theorem 3.5. *Let Assumptions 3.1 hold for system* (2)*. Suppose that* $(x_t^\epsilon, y_t^\epsilon)_{0\le t\le T}$ *and* $(X_t)_{0\le t\le T}$ *are realisations of the solution to* (2) *and* (16) *respectively, with* $(U_t, V_t)_{t\ge 0}$ *and* $(W_t)_{t\ge 0}$ *the corresponding realisation of the driving Brownian motion, where*

$$W_t = \tilde{q}^{-\frac{1}{2}} \left(\sqrt{q_1}U_t - a_{12}a_{22}^{-1}\sqrt{q_2}V_t\right), \tag{21}$$

for $\tilde{q}$ *defined in* (16)*. We also assume that* $x_0 = X_0$*. Then* $(x_t^\epsilon)_{0\le t\le T}$ *converges to* $(X_t)_{0\le t\le T}$ *in* L^2*. More specifically,*

$$\mathbb{E}\sup_{0\le t\le T}(x_t^\epsilon - X_t)^2 \le c\left(\epsilon^2\log(\frac{T}{\epsilon}) + \epsilon^2 T\right)e^{T^2}.$$

When T *is fixed and finite, the above bound will be of order* $\mathcal{O}(\epsilon^2\log(\epsilon))$*.*

Proof. We rewrite (2b) as

$$(a_{22}^{-1}a_{21}x_t^\epsilon + y_t^\epsilon)dt = \epsilon^2 a_{22}^{-1}dy_t^\epsilon - \epsilon a_{22}^{-1}\sqrt{q_2}dV_t\ . \tag{22}$$

We also rewrite (2a) as

$$\begin{aligned}
dx_t^\epsilon &= \frac{1}{\epsilon}a_{12}(a_{22}^{-1}a_{21}x_t^\epsilon + y_t^\epsilon)dt + a_{14}(a_{22}^{-1}a_{21}x_t^\epsilon + y_t^\epsilon)dt \\
&\quad +(a_{13} - a_{14}a_{22}^{-1}a_{21})x_t^\epsilon dt + \sqrt{q_1}dU_t \\
&= \left(\frac{1}{\epsilon}a_{12} + a_{14}\right)(a_{22}^{-1}a_{21}x_t^\epsilon + y_t^\epsilon)dt + \tilde{a}x_t^\epsilon dt + \sqrt{q_1}dU_t\ ,
\end{aligned}$$

where $\tilde{a}$ is defined in (17). Replacing $(a_{22}^{-1}a_{21}x_t^\epsilon + y_t^\epsilon)dt$ above by the right-hand-side of (22), we get

$$\begin{aligned}
dx_t^\epsilon &= \epsilon(a_{12} + \epsilon a_{14})a_{22}^{-1}dy_t^\epsilon - a_{12}a_{22}^{-1}\sqrt{q_2}dV_t - \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}dV_t \\
&\quad +\tilde{a}x_t^\epsilon dt + \sqrt{q_1}dU_t \\
&= \tilde{a}x_t^\epsilon dt + \epsilon(a_{12} + \epsilon a_{14})a_{22}^{-1}dy_t^\epsilon + \sqrt{\tilde{q}}dW_t - \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}dV_t\ .
\end{aligned}$$

Thus

$$\begin{aligned}
x_t^\epsilon &= x_0 + \int_0^t \tilde{a}x_s^\epsilon ds + \sqrt{\tilde{q}}W_t + \\
&\quad \epsilon(a_{12} + \epsilon a_{14})a_{22}^{-1}(y_t^\epsilon - y_0) - \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}V_t\ .
\end{aligned} \tag{23}$$

Recall that the solution to the homogenized equation (16) is given by

$$X_t = X_0 + \int_0^t \tilde{a}X_s ds + \sqrt{\tilde{q}}W_t\ . \tag{24}$$

Let $e(\epsilon)_t = x_t^\epsilon - X_t$. Subtracting the previous equation from (23) and using the assumption $X_0 = x_0$, we find that

$$e(\epsilon)_t = \tilde{a}\int_0^t e(\epsilon)_s ds + \epsilon\left((a_{12} + \epsilon a_{14})a_{22}^{-1}(y_t^\epsilon - y_0) - a_{14}a_{22}^{-1}\sqrt{q_2}V_t\right)\ .$$

Applying Lemma 3.4, we find an ϵ-independent constant C, such that

$$\mathbb{E}\left(\sup_{0\le t\le T}(y_t^\epsilon)^2\right) \le C\log(\frac{T}{\epsilon})\ .$$

By Cauchy-Schwarz,

$$\mathbb{E}\left(\sup_{0\leq t\leq T} e(\epsilon)_t^2\right) \leq c\left(T\int_0^T \mathbb{E}e(\epsilon)_s^2 ds + \epsilon^2\log(\frac{T}{\epsilon}) + \epsilon^2 T\right) . \tag{25}$$

By the integrated version of the Gronwall inequality [22], we deduce that

$$\mathbb{E}\left(\sup_{0\leq t\leq T} e(\epsilon)_t^2\right) \leq c\left(\epsilon^2\log(\frac{T}{\epsilon}) + \epsilon^2 T\right) e^{T^2} . \tag{26}$$

When T is finite, we have

$$\mathbb{E}\left(\sup_{0\leq t\leq T} e(\epsilon)_t^2\right) = \mathcal{O}\left(\epsilon^2\log(\epsilon)\right) .$$

This completes the proof. □

3.2 The Drift Estimator

As in the averaging case, a natural idea for estimating the drift of the homogenized equation is to use the maximum likelihood estimator (4), replacing X_t by the solution x_t^ϵ of (2a). However, in the case of homogenization we do not get asymptotically consistent estimates. To achieve this, we must subsample the data: we choose Δ (time step for observations) according to the value of the scale parameter ϵ and solve the estimation problem for discretely observed diffusions (see [21, 22, 23]). The maximum likelihood estimator for the drift of a homogenized equation converges after proper subsampling. We let the observation time interval Δ and the number of observations N both depend on the scaling parameter ϵ, by setting $\Delta = \epsilon^\alpha$ and $N = \epsilon^{-\gamma}$. We find the error is optimized in the L^2 sense when $\alpha = 1/2$. We will show that $\hat{a}_{N,\epsilon}$ converges to $\tilde{a}$ only if $\frac{\Delta}{\epsilon^2} \to \infty$, in a sense to be made precise later.

Theorem 3.6. *Suppose that $(x_t^\epsilon, y_t^\epsilon)_{t\geq 0}$ solves the system* (2) *satisfying Assumptions 3.1. Let $\hat{a}_{N,\epsilon}$ be the estimate we get by replacing X_t in* (4) *by x_t^ϵ and discretizing the integrals, i.e.*

$$\hat{a}_{N,\epsilon} = \left(\frac{1}{N\Delta}\sum_{n=0}^{N-1} x_n^\epsilon\left(x_{n+1}^\epsilon - x_n^\epsilon\right)\right)\left(\frac{1}{N\Delta}\sum_{n=0}^{N-1}(x_n^\epsilon)^2\Delta\right)^{-1} \tag{27}$$

Then,

$$\mathbb{E}(\hat{a}_{N,\epsilon} - \tilde{a})^2 = \mathcal{O}(\Delta^2 + \frac{1}{N\Delta} + \frac{\epsilon^2}{\Delta^2})$$

where $\tilde{a}$ as defined in (17). *Consequently, if $\Delta = \epsilon^\alpha$, $N = \epsilon^{-\gamma}$, $\alpha \in (0,1)$, $\gamma > \alpha$,*

$$\lim_{\epsilon\to 0}\mathbb{E}(\hat{a}_{N,\epsilon} - \tilde{a})^2 = 0 .$$

Furthermore, $\alpha = 1/2$ and $\gamma \geq 3/2$ optimize the error.

Before proving Theorem 3.6, we first note that the magnitude of the increment of y^ϵ over a small time interval Δ will be of order $\mathcal{O}(\frac{\sqrt{\Delta}}{\epsilon})$, coming from the discretization of the martingale part. By definition $\Delta = \epsilon^\alpha$. Thus, we conclude that

$$\mathbb{E}(y_{n+1}^\epsilon - y_n^\epsilon)^2 = \mathcal{O}(\epsilon^{\max(\alpha-2,0)}) , \tag{28}$$

taking into account the fact that, by ergodicity, this will never be more than $\mathcal{O}(1)$.

Proof. Define I_1 and I_2 as

$$I_1(\epsilon) = \frac{1}{N\Delta}\sum_{n=0}^{N-1}(x^\epsilon_{n+1} - x^\epsilon_n)x^\epsilon_n \ , \ I_2(\epsilon) = \frac{1}{N}\sum_{n=0}^{N-1}(x^\epsilon_n)^2$$

By ergodic theorem, and since $N = \epsilon^{-\gamma}$, we have

$$\lim_{\epsilon\to 0} I_2(\epsilon) = \mathbb{E}\left(X^2\right) = C \neq 0$$

which is a non-zero constant. Hence, it is sufficient to prove that

$$\mathbb{E}\left((I_1(\epsilon) - \tilde{a}I_2(\epsilon))^2\right) = \mathcal{O}(\Delta^2 + \frac{1}{N\Delta} + \frac{\epsilon^2}{\Delta^2}) \ .$$

We use the rearranged equation (23) of (2a) to decompose the error:

$$I_1(\epsilon) - \tilde{a}I_2(\epsilon) = J_1(\epsilon) + J_2(\epsilon) + J_3(\epsilon) + J_4(\epsilon) \ , \tag{29}$$

where

$$\begin{aligned}
J_1(\epsilon) &= \frac{\tilde{a}}{N\Delta}\sum_{n=0}^{N-1}\left(\int_{n\Delta}^{(n+1)\Delta} x^\epsilon_s ds - x^\epsilon_n\Delta\right)x^\epsilon_n, \\
J_2(\epsilon) &= \frac{1}{N\Delta}\sum_{n=0}^{N-1}\left(\sqrt{\tilde{q}}\int_{n\Delta}^{(n+1)\Delta} x^\epsilon_n dW_s\right), \\
J_3(\epsilon) &= \frac{\epsilon}{N\Delta}\sum_{n=0}^{N-1}(a_{12} + \epsilon a_{14})a_{22}^{-1}\int_{n\Delta}^{(n+1)\Delta} x^\epsilon_n dy^\epsilon_s, \\
J_4(\epsilon) &= \frac{\epsilon}{N\Delta}\sum_{n=0}^{N-1} a_{14}a_{22}^{-1}\sqrt{q_2}\int_{n\Delta}^{(n+1)\Delta} x^\epsilon_n dV_s.
\end{aligned}$$

By independence, Itô isometry and ergodicity, we immediately have

$$\begin{aligned}
\mathbb{E}\left(J_2(\epsilon)^2\right) &= \mathbb{E}\left(\frac{\sqrt{\tilde{q}}}{N\Delta}\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x^\epsilon_n dW_s\right)^2 \\
&= \frac{\tilde{q}}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x^\epsilon_n dW_s\right)^2 \\
&\leq \frac{\tilde{q}}{N^2\Delta^2}N\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta} dW_s\right)^2 \mathbb{E}\left((x^\epsilon_n)^2\right) \\
&\leq \frac{\tilde{q}}{N^2\Delta^2}N\Delta\mathbb{E}\left((x^\epsilon_n)^2\right) = \mathcal{O}(\frac{1}{N\Delta}) \ ,
\end{aligned}$$

and, similarly,

$$\mathbb{E}\left(J_4(\epsilon)^2\right) \leq \mathcal{O}(\frac{\epsilon^2}{N\Delta}) \ .$$

By Hölder inequality, and (28), we have,

$$\begin{aligned}
\mathbb{E}\left(J_3(\epsilon)^2\right) &= \mathbb{E}\left(\left(\frac{\epsilon C}{N\Delta}\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x_n^\epsilon dy^\epsilon\right)^2\right) = \mathbb{E}\left(\left(\frac{\epsilon C}{N\Delta}\sum_{n=0}^{N-1} x_n^\epsilon (y_{n+1}^\epsilon - y_n^\epsilon)\right)^2\right) \\
&\leq \frac{\epsilon^2}{N^2\Delta^2}\mathbb{E}\left(\left(\sum_{n=0}^{N-1}(y_{n+1}^\epsilon - y_n^\epsilon)\right)^2\right)\mathbb{E}\left((\sum_{n=0}^{N-1} x_n^\epsilon)^2\right) \\
&\leq \frac{\epsilon^2 C}{N^2\Delta^2} N(\epsilon^{\max(\alpha-2,0)}) N\mathbb{E}\left((x_n^\epsilon)^2\right) = \mathcal{O}(\frac{\epsilon^2}{\Delta^2})\ .
\end{aligned}$$

It remains to get an estimate for $J_1(\epsilon)$. We use the integrated form of equation (23) on time interval $[n\Delta, s]$ to replace x_s^ϵ

$$\mathbb{E}\left(J_1(\epsilon)^2\right) = \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\left(\left(\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}(x_s^\epsilon - x_n^\epsilon)x_n^\epsilon ds\right)^2\right) \tag{30}$$

$$= \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}(K_1^{(n,\epsilon)} + K_2^{(n,\epsilon)} + K_3^{(n,\epsilon)} + K_4^{(n,\epsilon)})\right)^2 \tag{31}$$

(32)

where,

$$\begin{aligned}
K_1^{(n,\epsilon)} &= \tilde{a}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_{n\Delta}^\epsilon x_u^\epsilon du ds\ , \\
K_2^{(n,\epsilon)} &= \epsilon(a_{12} + \epsilon a_{14})a_{22}^{-1}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_{n\Delta}^\epsilon dy_u^\epsilon ds\ , \\
K_3^{(n,\epsilon)} &= \sqrt{\tilde{q}}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_{n\Delta}^\epsilon dW_u ds\ , \\
K_4^{(n,\epsilon)} &= \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_{n\Delta}^\epsilon dV_u ds\ .
\end{aligned}$$

We immediately see that

$$\mathbb{E}\left(J_1(\epsilon)^2\right) = \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}\left(\sum_{i=1}^{4}K_i^{(n,\epsilon)}\right)^2\right) \tag{33}$$

$$+ \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\left(\sum_{m\neq n}\left(\sum_{i=1}^{4}K_i^{(n,\epsilon)}\right)\left(\sum_{j=1}^{4}K_j^{(m,\epsilon)}\right)\right) \tag{34}$$

Remark 3.7. *Under the vector valued problem, we use the exact decomposition of $\mathbb{E}\|J_1(\epsilon)\|^2$ by using* (33) *and* (34)*. This is essential in order to obtain more optimized subsampling rate for the drift estimator. For general L^p bound for the error, Holder's inequality leads to an optimal subsampling rate of $\alpha = 2/3$, and achieves an over all L^1 error of order $\mathcal{O}(\epsilon^{1/3})$ [23]. However, this magnitude of overall error is not optimal in L^2. We will show later that the optimal L^2 error can be achieved at the order of $\mathcal{O}(\epsilon^{1/2})$, using the exact decomposition shown above.*

By Cauchy-Schwarz inequality, we know for line (33),

$$\mathbb{E}\left(\sum_{n=0}^{N-1}\left(\sum_{i=1}^{4} K_i^{(n,\epsilon)}\right)^2\right) \leq \sum_{n=0}^{N-1}\sum_{i=1}^{4}\mathbb{E}\left(\left(K_i^{(n,\epsilon)}\right)^2\right) .$$

Using first order iterated integrals, we have

$$\begin{aligned}
\mathbb{E}\left((K_1^{(n,\epsilon)})^2\right) &= \mathbb{E}\left(\left(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_{n\Delta}^{\epsilon} x_u^{\epsilon} du ds\right)^2\right) \\
&\leq C\Delta\int_{n\Delta}^{(n+1)\Delta}\mathbb{E}\left(\int_{n\Delta}^{s}(x_u^{\epsilon})^2 du ds (x_{n\Delta}^{\epsilon})^2\right) \\
&\leq C\Delta\int_{n\Delta}^{(n+1)\Delta}(s-n\Delta)^2 ds \\
&= \mathcal{O}(\Delta^4) .
\end{aligned}$$

Using (28), we have

$$\begin{aligned}
\mathbb{E}\left((K_2^{(n,\epsilon)})^2\right) &= \mathbb{E}\left(\epsilon C\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_{n\Delta}^{\epsilon} dy_u^{\epsilon} ds\right) \\
&\leq C\epsilon^2\mathbb{E}\left(\left(\int_{n\Delta}^{(n+1)\Delta} x_{n\Delta}^{\epsilon}(y_s^{\epsilon}-y_u^{\epsilon}) ds\right)^2\right) \\
&\leq C\epsilon^2\Delta\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}(y_s^{\epsilon}-y_u^{\epsilon})^2 ds (x_{n\Delta}^{\epsilon})^2\right) \\
&\leq C\epsilon^2\Delta\int_{n\Delta}^{(n+1)\Delta}(e^{-\frac{s-n\Delta}{\epsilon^2}}-1) ds \\
&= \mathcal{O}\left(\epsilon^4(e^{-\frac{\Delta}{\epsilon^2}}-1)\right) .
\end{aligned}$$

For $K_3^{(n,\epsilon)}$, we have,

$$\begin{aligned}
\mathbb{E}\left((K_3^{(n,\epsilon)})^2\right) &= \mathbb{E}\left(\left(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s}\sqrt{\tilde{q}} x_{n\Delta}^{\epsilon} dW_u ds\right)^2\right) \\
&\leq C\Delta\int_{n\Delta}^{(n+1)\Delta}\mathbb{E}\left(\left(\int_{n\Delta}^{s} dW_u\right)^2\right) ds \\
&\leq C\Delta\int_{n\Delta}^{(n+1)\Delta}(s-n\Delta) ds \\
&= \mathcal{O}(\Delta^3) .
\end{aligned}$$

Since $K_4^{(n,\epsilon)}$ is similar to $K_3^{(n,\epsilon)}$, we have

$$\mathbb{E}\left((K_4^{(n,\epsilon)})^2\right) = \mathcal{O}(\epsilon^2\Delta^3) .$$

Thus, for line (33), the order of the dominating terms are,

$$\mathbb{E}\left(\sum_{n=0}^{N-1}\left(\sum_{i=1}^{4}K_i^{(n,\epsilon)}\right)^2\right)=\mathcal{O}(N\Delta^4+N\epsilon^4(e^{-\frac{\Delta}{\epsilon^2}}-1)+N\Delta^3)\ .$$

For line (34),

$$\mathbb{E}\left(\sum_{m\neq n}(\sum_{i=1}^{4}K_i^{(n,\epsilon)})(\sum_{j=1}^{4}K_j^{(m,\epsilon)})\right)\leq\sum_{m\neq n}\mathbb{E}\left(\sum_{i=1}^{4}K_i^{(n,\epsilon)}\right)\mathbb{E}\left(\sum_{j=1}^{4}K_j^{(m,\epsilon)}\right)\ .$$

We know,

$$\begin{aligned}\mathbb{E}(K_1^{(n,\epsilon)}) &= \mathbb{E}\left(C\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s}x_u^\epsilon du ds\right)\\ &\leq C\left(\int_{n\Delta}^{(n+1)\Delta}(s-n\Delta)ds\right)\\ &= \mathcal{O}(\Delta^2)\ .\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathbb{E}\left(K_2^{(n,\epsilon)}\right) &= \epsilon C\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}(y_s^\epsilon-y_{n\Delta}^\epsilon)ds\right)\\ &= \mathcal{O}(\epsilon\Delta)\ .\end{aligned}$$

Since the integral of Brownian motions is Gaussian

$$\begin{aligned}\mathbb{E}\left(K_3^{(n,\epsilon)}\right) &= C\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s}dW_u ds)\\ &= C\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}(W_s-W_{n\Delta})ds)\\ &= C\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}W_s ds-W_{n\Delta}\Delta)=0\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left(K_4^{(n,\epsilon)}\right) &= C\epsilon\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s}dV_u ds)\\ &= C\epsilon\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}V_s ds-V_{n\Delta}\Delta)=0.\end{aligned}$$

Thus,

$$\mathbb{E}\left(\sum_{i=1}^{4}K_i^{(n,\epsilon)}\right)=\mathcal{O}(\Delta^2+\epsilon\Delta)\ ,$$

immediately we have for line (34),

$$\mathbb{E}\left(\sum_{m\neq n}\left((\sum_{i=1}^{4}K_i^{(n,\epsilon)})(\sum_{j=1}^{4}K_j^{(m,\epsilon)})\right)\right)=\mathcal{O}(N^2\Delta^4+N^2\epsilon^2\Delta^2)\ .$$

Putting all terms for J_1 together, we keep the dominating terms, and by assumption $N\Delta \to \infty$, and $\alpha < 2$ since $e^{-\frac{\Delta}{\epsilon^2}} \to 0$,

$$\begin{aligned}
\mathbb{E}\left(J_1(\epsilon)^2\right) &\leq \frac{C}{N^2\Delta^2}(N\Delta^4 + N\epsilon^4(e^{-\frac{\Delta}{\epsilon^2}} - 1) + N\Delta^3) \\
&+ \frac{C}{N^2\Delta^2}(N^2\Delta^4 + N^2\epsilon^2\Delta^2) \\
&= \mathcal{O}(\frac{\Delta^2}{N} + \frac{\epsilon^4}{N\Delta^2}(e^{-\frac{\Delta}{\epsilon^2}} - 1) + \frac{\Delta}{N} + \Delta^2 + \epsilon^2) \\
&= \mathcal{O}(\frac{\epsilon^4}{N\Delta^2} + \Delta^2 + \epsilon^2)\ .
\end{aligned}$$

Therefore, putting $J_i(\epsilon)$'s, $i \in \{1, 2, 3, 4\}$, together, we have,

$$\begin{aligned}
\mathbb{E}\left((I_1(\epsilon) - \tilde{a}I_2(\epsilon))^2\right) &\leq \sum_{i=1}^{4} \mathbb{E}\left(J(\epsilon)_i^2\right) \\
&= \mathcal{O}(\frac{\epsilon^4}{N\Delta^2} + \Delta^2 + \epsilon^2) \\
&\quad + \mathcal{O}(\frac{1}{N\Delta}) + \mathcal{O}(\frac{\epsilon^2}{\Delta^2}) + \mathcal{O}(\frac{\epsilon^2}{N\Delta}) \\
&= \mathcal{O}(\Delta^2 + \frac{1}{N\Delta} + \frac{\epsilon^2}{\Delta^2})
\end{aligned}$$

We rewrite the above equation using $\Delta = \epsilon^\alpha$ and $N = \epsilon^{-\gamma}$,

$$\mathbb{E}\left((I_1(\epsilon) - \tilde{a}I_2(\epsilon))^2\right) = \mathcal{O}(\epsilon^{2\alpha} + \epsilon^{\gamma-\alpha} + \epsilon^{2-2\alpha})\ .$$

It is immediately seen that $\alpha = \frac{1}{2}$ and $\gamma \geq 3/2$ optimize the error, and $\alpha \in (0, 1)$, the order of the error is

$$\mathbb{E}\left((I_1(\epsilon) - \tilde{a}I_2(\epsilon))^2\right) = \mathcal{O}(\epsilon)\ .$$

This completes the proof. □

In Figure 3, we show an example of the L^2 error of the drift estimator with various scaling parameter ϵ and subsampling rate α. We see that the error is minimized around $\alpha = 1/2$ as in Theorem 3.6.

3.3 The Diffusion Estimator

Just as in the case of the drift estimator, we define the diffusion estimator by the maximum likelihood estimator (5), where X is replaced by the discretized solution of (2a). More specifically, we define

$$\tilde{q}^{\epsilon}_{N,\Delta} = \frac{1}{N\Delta}\sum_{n=0}^{N-1}(x^{\epsilon}_{n+1} - x^{\epsilon}_{n})^2 \tag{35}$$

where $x^{\epsilon}_{n} = x^{\epsilon}_{n\Delta}$ is the discrete observation of the process generated by (2a) and Δ is the observation time interval.

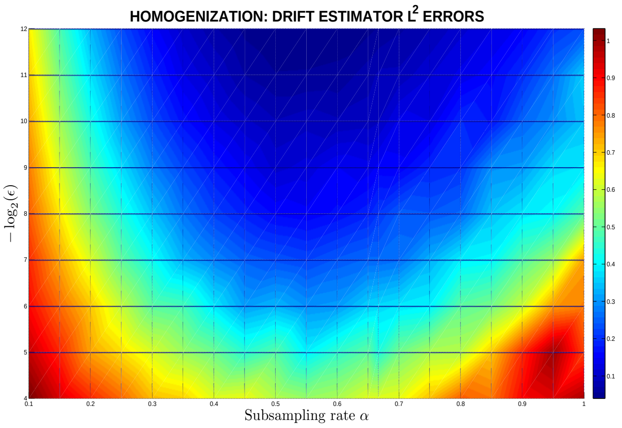

Figure 3: This is a colormap plot of the L^2 norm of the errors from the drift estimator $\hat{a}_{N,\epsilon}$ at different subsampling rates. Simulations are done through exact solution of the multiscale OU system. Each path is subsampled with $N = \epsilon^{1.5}$ number of observations, at time increment of $\Delta = \epsilon^{\alpha}$, with $\alpha \in [0.1, 1]$. We take ϵ from choices of 2^{-4} to 2^{-12}. Each estimate is based on 100 paths. The initial condition is $(x_0, y_0) = (0, 0)$ and the parameter values are $a_{11} = a_{12} = a_{13} = a_{21} = a_{22} = -1$, $a_{14} = 1$, $q_1 = q_2 = 2$.

Theorem 3.8. *Suppose that $(x_t^{\epsilon})_{t>0}$ is the projection to the x-coordinate of a solution of system* (2) *satisfying Assumptions 3.1. Let $\hat{q}_{\epsilon}$ be the estimate we get by replacing X in* (5) *by x^{ϵ}, i.e.*

$$\hat{q}_{\epsilon} = \frac{1}{T} \sum_{n=0}^{N-1} (x_{n+1}^{\epsilon} - x_n^{\epsilon})^2 \ .$$

Then

$$\mathbb{E}\left((\hat{q}_{\epsilon} - \tilde{q})^2\right) = \mathcal{O}\left(\Delta + \epsilon^2 + \frac{\epsilon^4}{\Delta^2}\right)$$

where $\tilde{q}$ as defined in (18)*. Consequently, if $\Delta = \epsilon^{\alpha}$, fix $T = N\Delta$, and $\alpha \in (0, 2)$, then*

$$\lim_{\epsilon \to 0} \mathbb{E}\left((\hat{q}_{\epsilon} - \tilde{q})^2\right) = 0 \ .$$

Furthermore, $\alpha = 4/3$ optimizes the error.

We first define

$$\sqrt{\Delta}\eta_n = \int_{n\Delta}^{(n+1)\Delta} dW_t \ .$$

Proof. We now prove Theorem 3.8. Using the integral form of equation (23),

$$\begin{aligned} x_{n+1}^{\epsilon} - x_n^{\epsilon} &= \int_{n\Delta}^{(n+1)\Delta} \sqrt{\tilde{q}} dW_s \\ &+ \hat{R}_{1,\epsilon} + \hat{R}_{2,\epsilon} + \hat{R}_{3,\epsilon} \end{aligned} \tag{36}$$

where

$$
\begin{aligned}
\hat{R}_{1,\epsilon} &= \tilde{a}\int_{n\Delta}^{(n+1)\Delta} x_s^\epsilon ds \\
\hat{R}_{2,\epsilon} &= \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}\int_{n\Delta}^{(n+1)\Delta} dV_s \\
\hat{R}_{3,\epsilon} &= \epsilon(a_{12}+\epsilon a_{14})a_{22}^{-1}\int_{n\Delta}^{(n+1)\Delta} dy(s)
\end{aligned}
$$

We rewrite line (36) as

$$
\int_{n\Delta}^{(n+1)\Delta} \sqrt{\tilde{q}}dW_s = \sqrt{\tilde{q}\Delta}\eta_n \tag{37}
$$

where η_n are $\mathcal{N}(0,1)$ random variables.

For Δ and ϵ sufficiently small, by Cauchy-Schwarz inequality

$$
\begin{aligned}
\mathbb{E}\left(\left(c\int_{n\Delta}^{(n+1)\Delta} x_s^\epsilon ds\right)^2\right) &\le c\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}(x_s^\epsilon)^2 ds\int_{n\Delta}^{(n+1)\Delta} ds\right) \\
&\le c\Delta\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}(x_s^\epsilon)^2 ds\right) \\
&\le c\Delta^2\mathbb{E}\left(\sup_{n\Delta\le s\le(n+1)\Delta}(x_s^\epsilon)^2\right) \\
&= \mathcal{O}(\Delta^2)
\end{aligned}
$$

Therefore,

$$
\mathbb{E}\left((\hat{R}_{1,\epsilon})^2\right) = \mathcal{O}(\Delta^2)
$$

By Itô isometry

$$
\mathbb{E}\left((\hat{R}_{2,\epsilon})^2\right) = \mathcal{O}(\epsilon^2\Delta)
$$

Then we look at $\hat{R}_{3,\epsilon}$,

$$
\mathbb{E}\left((\hat{R}_{3,\epsilon})^2\right) = \epsilon^2 C\mathbb{E}\left((y_{n+1}^\epsilon - y_n^\epsilon)^2\right)
$$

By (28), we have

$$
\mathbb{E}\left((\hat{R}_{3,\epsilon})^2\right) = \mathcal{O}(\epsilon^{\max(\alpha,2)}) \tag{38}
$$

We substitute $(x_{n+1}^\epsilon - x_n^\epsilon)$ into the estimator $\hat{q}_\epsilon$ in Theorem 3.8. We decompose the estimator's error as follows,

$$\begin{aligned}
\hat{q}_\epsilon - \tilde{q} &= \tilde{q}(\frac{1}{N}\sum_{n=0}^{N-1}\eta_n^2 - 1) \\
&+ \frac{1}{T}\sum_{n=0}^{N-1}\sum_{i=1}^{3}\left(\hat{R}_{i,\epsilon}^2\right) \\
&+ \frac{2}{T}\sum_{n=0}^{N-1}\sum_{i=1}^{3}\hat{R}_{i,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n \\
&+ \frac{1}{T}\sum_{n=0}^{N-1}\left(\sum_{i\neq j}\hat{R}_{i,\epsilon}\hat{R}_{j,\epsilon}\right) \\
&= R_\epsilon
\end{aligned}$$

Then we bound the mean squared error using Cauchy-Schwarz inequality.

$$\begin{aligned}
\mathbb{E}\left((\hat{q}_\epsilon - \tilde{q})^2\right) &\leq C\tilde{q}^2\mathbb{E}\left((\frac{1}{N}\sum_{n=0}^{N-1}\eta_n^2 - 1)^2\right) && (39) \\
&+ C\sum_{i=1}^{3}\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{i,\epsilon}^2\right)^2\right) && (40) \\
&+ C\sum_{i=1}^{3}\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{i,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right) && (41) \\
&+ C\sum_{i\neq j}\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\left(\hat{R}_{i,\epsilon}\otimes\hat{R}_{j,\epsilon}\right)\right)^2\right) && (42)
\end{aligned}$$

By law of large numbers, line (39) is of order $\mathcal{O}(\Delta)$.

In line (40), for $i \in \{1, 2\}$, we have

$$\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{i,\epsilon}^2\right)^2\right) = \frac{1}{T^2}N\sum_{n=0}^{N-1}\mathbb{E}\left((\hat{R}_{i,\epsilon}^2)^2\right).$$

Since $\mathbb{E}\left((\hat{R}_{1,\epsilon})^2\right) = \mathcal{O}(\Delta^2)$, we have

$$\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{1,\epsilon}^2\right)^2\right) = \mathcal{O}\left(N^2(\Delta^2)^2\right) = \mathcal{O}\left(\Delta^2\right);$$

since $\mathbb{E}\left((\hat{R}_{2,\epsilon})^2\right) = \mathcal{O}(\epsilon^2\Delta)$, we have

$$\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{2,\epsilon}^2\right)^2\right) = \mathcal{O}\left(N^2(\Delta\epsilon^2)^2\right) = \mathcal{O}(\epsilon^4).$$

The estimate is different for $\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{3,\epsilon}^2\right)^2\right)$. By (28), we have

$$\begin{aligned}\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{3,\epsilon}^2\right)^2\right) &= \frac{C\epsilon^4}{T^2}\mathbb{E}\left(\left(\sum_{n=0}^{N-1}\left(y_{n+1}^\epsilon - y_n^\epsilon\right)^2\right)^2\right)\\ &\le C\epsilon^4 N\sum_{n=0}^{N-1}\mathbb{E}\left(\left(y_{n+1}^\epsilon - y_n^\epsilon\right)^4\right)\\ &= \mathcal{O}\left(\frac{\epsilon^{4+2\max(0,\alpha-2)}}{\Delta^2}\right)\\ &= \mathcal{O}\left(\frac{\epsilon^{\max(4,2\alpha)}}{\Delta^2}\right)\end{aligned}$$

Adding up all terms for line (40), we have,

$$\sum_{i=1}^{3}\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{i,\epsilon}^2\right)^2\right) = \mathcal{O}\left(\Delta^2 + \epsilon^4 + \frac{\epsilon^{\max(4,2\alpha)}}{\Delta^2}\right). \tag{43}$$

In line (41), for $i \in \{1,2\}$, we have

$$\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{i,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right) \le CN^2\Delta\mathbb{E}\left(\left(\hat{R}_{i,\epsilon}\eta_n\right)^2\right) = CN\mathbb{E}\left((\hat{R}_{i,\epsilon})^2\right)$$

Since $\mathbb{E}\left((\hat{R}_{1,\epsilon})^2\right) = \mathcal{O}(\Delta^2)$, we have

$$\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{1,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right) = \mathcal{O}(N\Delta^2) = \mathcal{O}(\Delta);$$

since $\mathbb{E}\left((\hat{R}_{2,\epsilon})^2\right) = \mathcal{O}(\epsilon^2\Delta)$, we have

$$\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{2,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right) = \mathcal{O}(N\epsilon^2\Delta) = \mathcal{O}(\epsilon^2)\ .$$

Again, it is different for $\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{3,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right)$ due to correlation between $\hat{R}_{3,\epsilon}^{(n)}$ and η_n. Using the expression from (37) by only considering the dominating terms, we have

$$\begin{aligned}&\mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{3,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right)\\ =\ &\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{3,\epsilon}^2\left(\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right)\\ +\ &\mathbb{E}\left(\frac{1}{T^2}\sum_{m\neq n}\hat{R}_{3,\epsilon}^{(m)}\hat{R}_{3,\epsilon}^{(n)}\int_{m\Delta}^{(m+1)\Delta}\sqrt{\tilde{q}}dWs\int_{n\Delta}^{(n+1)\Delta}\sqrt{\tilde{q}}dWs\right)\end{aligned}$$

By computing the order of the dominating terms and the martingale terms, when $m = n$,

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_{3,\epsilon}^{2}\left(\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right) &= \frac{1}{T}\sum_{n=0}^{N-1}\Delta\mathbb{E}\left(\hat{R}_{3,\epsilon}^{2}\tilde{q}\eta_n^2\right)\\
&= \frac{1}{T}\mathbb{E}(\hat{R}_{3,\epsilon}^{2}\eta_n^2)\\
&= \mathcal{O}\left(\epsilon^{\max(\alpha,2)}\right)
\end{aligned}$$

and when $m < n$,

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{T^2}\sum_{m\neq n}\hat{R}_{3,\epsilon}^{(m)}\hat{R}_{3,\epsilon}^{(n)}\int_{m\Delta}^{(m+1)\Delta}\sqrt{\tilde{q}}dWs\int_{n\Delta}^{(n+1)\Delta}\sqrt{\tilde{q}}dWs\right)\\
\leq\;& CN^2\epsilon^2\mathbb{E}\left((y_{n+1}^{\epsilon}-y_n^{\epsilon})(y_{m+1}^{\epsilon}-y_m^{\epsilon})\right.\\
\times\;& \left.\int_{n\Delta}^{(n+1)\Delta}dW_s\int_{m\Delta}^{(m+1)\Delta}dW_s\right)\\
\leq\;& CN^2\epsilon^2\mathbb{E}\left((y_{n+1}^{\epsilon}-y_n^{\epsilon})\int_{n\Delta}^{(n+1)\Delta}dW_s\right.\\
\times\;& \left.\mathbb{E}\left((y_{m+1}^{\epsilon}-y_m^{\epsilon})\int_{m\Delta}^{(m+1)\Delta}dW_s|\mathcal{F}_{m\Delta}\right)\right)
\end{aligned}$$

Using the expansion in (37), and using the dominating terms only,

$$\begin{aligned}
&\mathbb{E}\left((y_{m+1}^{\epsilon}-y_m^{\epsilon})\int_{n\Delta}^{(n+1)\Delta}dW_s|\mathcal{F}_{m\Delta}\right)\\
=\;& \mathbb{E}\left(\left((e^{-\frac{\Delta}{\epsilon^2}}-1)y_m^{\epsilon}\right.\right.\\
+\;& \frac{1}{\epsilon^2}\int_{m\Delta}^{(m+1)\Delta}e^{-\frac{(m+1)\Delta-s}{\epsilon^2}}x_s^{\epsilon}ds\\
+\;& \left.\left.\frac{1}{\epsilon}\int_{m\Delta}^{(m+1)\Delta}e^{-\frac{(m+1)\Delta-s}{\epsilon^2}}dV_s\right)\int_{m\Delta}^{(m+1)\Delta}dW_s|\mathcal{F}_{m\Delta}\right)\\
=\;& \mathcal{O}(\epsilon(e^{-\frac{\Delta}{\epsilon^2}}-1))
\end{aligned}$$

Therefore, when $m < n$, we have,

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{T^2}\sum_{m\neq n}\hat{R}_{3,\epsilon}^{(m)}\hat{R}_{3,\epsilon}^{(n)}\int_{m\Delta}^{(m+1)\Delta}\sqrt{\tilde{q}}dWs\int_{n\Delta}^{(n+1)\Delta}\sqrt{\tilde{q}}dWs\right)\\
=\;& \mathcal{O}\left(\frac{\epsilon^4}{\Delta^2}(e^{-\frac{\Delta}{\epsilon^2}}-1)^2\right)\\
=\;& \mathcal{O}\left(\epsilon^{4-2\alpha+2\max(\alpha-2,0)}\right)\\
=\;& \mathcal{O}\left(\epsilon^{\max(0,4-2\alpha)}\right)
\end{aligned}$$

In the case $m > n$, the result is identical due to symmetry. Adding up all terms for line (41),

$$\sum_{i=1}^{5} \mathbb{E}\left(\left(\frac{1}{T}\sum_{n=0}^{N-1} \hat{R}_{i,\epsilon}\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right)$$
$$= \mathcal{O}\left(\Delta + \epsilon^2 + \epsilon^{\max(\alpha,2)} + \epsilon^{2\max(0,2-\alpha)}\right) \tag{44}$$

In line (42), we have

$$\sum_{i\neq j} \mathbb{E}\left(\left(\sum_{n=0}^{N-1} \hat{R}_{i,\epsilon}\hat{R}_{j,\epsilon}\right)^2\right) \leq N\mathbb{E}\left((R_{i,\epsilon})^2 \mathbb{E}(R_{j,\epsilon})^2\right)$$

Substituting in the L^2 norms of each $\hat{R}_{i,\epsilon}$, $i \in \{1, 2, 3\}$, we have for line (42),

$$\sum_{i\neq j} \mathbb{E}\left(\left(\sum_{n=0}^{N-1} \hat{R}_{i,\epsilon}\hat{R}_{j,\epsilon}\right)^2\right)$$
$$= \mathcal{O}\left(\Delta^2\epsilon^2 + \Delta\epsilon^{\max(\alpha,2)} + \epsilon^{2+\max(\alpha,2)}\right) \tag{45}$$

Aggregating bounds (43), (44) and (45) for equation lines from (39) to (42) respectively, we have

$$\begin{aligned} & \mathbb{E}\left((\hat{q}_\epsilon - \tilde{q})^2\right) \\ = \; & \mathcal{O}(\Delta) \\ + \; & \mathcal{O}\left(\Delta^2 + \epsilon^4 + \frac{\epsilon^{\max(4,2\alpha)}}{\Delta^2}\right) \\ + \; & \mathcal{O}\left(\Delta + \epsilon^2 + \epsilon^{\max(\alpha,2)} + \epsilon^{2\max(0,2-\alpha)}\right) \\ + \; & \left(\Delta^2\epsilon^2 + \Delta\epsilon^{\max(\alpha,2)} + \epsilon^{2+\max(\alpha,2)}\right) \end{aligned}$$

It is clear that when $\alpha < 2$,

$$\mathbb{E}\left((\hat{q}_\epsilon - \tilde{q})^2\right) = \mathcal{O}(\Delta + \epsilon^{4-2\alpha} + \epsilon^2).$$

The error is minimized when $\alpha = 4/3$, which is of order

$$\mathbb{E}\left((\hat{q}_\epsilon - \tilde{q})^2\right) = \mathcal{O}\left(\epsilon^{\frac{4}{3}}\right) .$$

It is easy to see when $\alpha > 2$, the error explodes. This completes the proof. $\square$

In Figure 4, we show an example of the L^2 error of the diffusion parameter with various scaling parameter ϵ and subsampling rate α. We see that the error is minimized around $\alpha = 4/3$ as in Theorem 3.8.

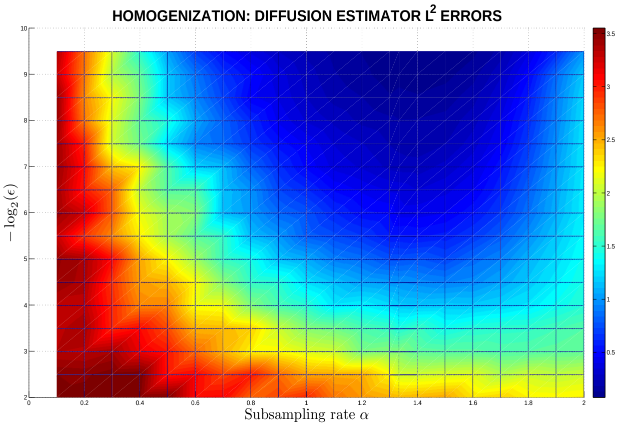

Figure 4: This is a colormap of the L^2 norm of $(\hat{q}_\epsilon - \tilde{q})$ for different ϵ and α. Each path is generated over a fixed total time horizon of $T = 1$, at a very fine resolution with $\delta = 2^{-20}$, with available number of observations $N = 2^{20}$. Each estimate is based on 100 paths. We test the scale parameter ϵ from 2^{-2} to $2^{-9.5}$, and test the diffusion estimator at a sequence of subsampling rates α over each path at rates $[0.1, 2]$. The system's parameter values are $a_{11} = a_{12} = a_{13} = a_{21} = a_{22} = -1$, $a_{14} = 1$, $q_1 = q_2 = 2$.

References

[1] Aït-Sahalia, Y.; Mykland, P. A; Zhang, L. A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data *Journal of the American Statistical Association* **100(472)** 2005

[2] Aït-Sahalia, Y.; Mykland, P. A; Zhang, L. How Often To Sample a Continuous-Time Process in the Presence of Market Microstructure Noise *The Review of Financial Studies* **18** 2005

[3] Azencott, R.; Beri A.; Timofeyev, I. Adaptive sub-sampling for parametric estimation of gaussian diffusions. *Journal of Statistical Physics* **139(6)**, 2010

[4] Azencott, R.; Beri A.; Timofeyev, I. Sub-sampling and parametric estimation for multiscale dynamics. *SIAM Multiscale Model. Simul.* 2010

[5] Anderson, D. F.; Cracium, G.; Kurtz, T. Product-form Stationary Distributions for Deficiency Zero Chemical Reaction Networks. *Bulletin of mathematical biology* **72(8)**, 2010

[6] Bishwal, J.P.N. Parameter Estimation in Stochastic Differential Equations Springer-Verlag 2008

[7] Chen, X. Limit theorems for functionals of ergodic Markov Chains with general state space. *Memoirs of the AMS* **129(664)**, 1999

[8] Crisan, D. Stochastic Analysis Springer, 2010

[9] El Machkouri, M., Es-Sebaiy, K. and Ouknine, Y.. Least squares estimator for non-ergodic OrnsteinUhlenbeck processes driven by Gaussian processes. *Journal of the Korean Statistical Society* **45**(3): pp 329–341, 2016.

[10] Graversen, S. E.; Peskir, G. Maximal Inequalities for the Ornstein-Uhlenbeck Process *Proceedings of the American Mathematical Society* **128(10)**, 2000

[11] Hall, P. and Heyde, C.C Martingale limit theory and its application. Academic Press, 1980.

[12] Isaacson, E.; Keller, H. B. Analysis of Numerical Methods Wiley & Sons. 1966.

[13] Ibragimov, I.; Hasminskii, R. Statistical estimation: Asymptotic theory. Springer, 1979.

[14] Kabanov, Y.; Pergamenschikov, S. Two-scale stochastic system: Asymptotic analysis and control. Springer, 2003.

[15] Kleptsyna, M. L. and Le Breton, A.. Statistical analysis of the fractional OrnsteinUhlenbeck type process. *Statistical Inference for stochastic processes* **5**(3): 229–248, 2002.

[16] Kutoyants, Y. A. Statistical inference for ergodic diffusion processes. Springer-Verlag, 2004.

[17] Liu, D.; Strong convergence of principle of averaging for multiscale stochastic dynamcial systems *Comm. Math. Sci* **8(4)** 2010

[18] Maslowski, B. and Tudor, C.A.. Drift parameter estimation for infinite-dimensional fractional OrnsteinUhlenbeck process. *Bulletin des Sciences Mathmatiques* **137**(7): 880–901, 2013.

[19] Miao, W. Estimation of diffusion parameters in diffusion processes and their asymptotic normality *International Journal of Comtemporary Mathematical Science* **1(16)** 2006

[20] Olhede, S; Pavliotis, G. A. ; Sykulski, A. Multiscale Inference for High Frequency Data *Tech. Rep. 290 Department of Statistical Science, UCL* 2008

[21] Papavasiliou, A.; Pavlioitis, G. A.; Stuart, A. M. Maximum Likelihood Drift Estimation for Multiscale Diffusions *Stoch. Proc. Appl.* **119(10)** 2009

2003

[22] Pavliotis, G. A. ; Stuart, A. M. Multiscale Methods Averaging and Homogenization. Springer, 2008.

[23] Pavliotis, G. A. ; Stuart, A. M. Parameter Estimation for Multiscale Diffusions *J. Stat. Phys.* **127** 2007

[24] Prakasa Rao, B. L. S. Semimartingales and their statistical inference. *Monograph on Statistics and Applied Probability*, **83** Chapman & Hall/CRC, 1999